ROBUST CEPHALOMETRIC LANDMARK IDENTIFICATION USING SUPPORT VECTOR MACHINES

Shantanu Chakrabartty[1*], *Masakazu Yagi*[2], *Tadashi Shibata*[3] *and Gert Cauwenberghs*[1]

[1] Center for Language and Speech Processing,
The Johns Hopkins University, Baltimore, MD 21218, USA
[2] Department of Electronics,
[3] Department of Frontier Informatics,
University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan

ABSTRACT

A robust and accurate image recognizer for cephalometric landmarking is presented. The recognizer uses $Gini$ Support Vector Machine (SVM) to model discrimination boundaries between different landmarks and also between the background frames. Large Margin Classification with non-linear kernels allows to extract relevant details from the landmarks, approaching human expert levels of recognition. In conjunction with Projected Principal-Edge Distribution (PPED) representation as feature vectors, $Gini$SVM is able to demonstrate more than 95% accuracy for landmark detection on medical cephalograms within a reasonable location tolerance value.

1. INTRODUCTION

In Cephalometric landmark identification, dentists are required to identify predefined characteristic anatomical landmarks on a cephalometric radio-graph (x-ray head film) to diagnose cranial bone structures of their patients. A sample head film with eight distinct landmarks is shown in Figure 1. The difficulty of identifying these landmarks is compounded by variability of patients skull structure and nature of the radio-graph image, for which most dentists have to use their expertise gained through several years of clinical practice. Any competitive image recognition system has to match the accuracy close to human performance, and thus stringent requirement on tolerances of location estimation are imposed. For a typical orthodontic application a reasonable tolerance distance for identification is around 1mm, which amounts to a resolution of about 4 pixels relative to the size of the images chosen in this work. In the Figure 1 a 5mm tolerance boundary around the landmark is shown.

*This work was performed while the author was at University of Tokyo, through an Academic Frontiers Student Exchange Promotion Program Scholarship and a fellowship from The Catalyst Foundation, New York.

Projected Principal-Edge Distribution (PPED), previously called Principal Axis Projection (PAP) is an effective feature extraction tool that has been designed for such a task [2]. The elegance of this method lies in its simplicity that enables real-time image recognition using dedicated hardware. In [2] [3] an associative memory back-end was employed to detect the landmark positions in medical radiographs. Like most maximum likelihood techniques such a method requires complicated parametric models and hence large amount of training data to reliably model class distributions. The performance of such classifiers deteriorates further as the degree of class overlap increases. If the aim of a recognizer were to discriminate between classes, it would suffice to model the decision boundaries which in most affine cases require much fewer parameters to estimate.

Large Margin Classifier like Support Vector Machines (SVMs) are one such classifier that is an attractive choice for implementing a back-end of an image recognition system because

1. They generalize well even with relatively few data points in the training set and bound on the generalization error can be directly estimated from the training data.
2. The only parameter that needs to be chosen is a penalty term for classification which acts as a regularizer and determines a trade-off between resolution and generalization performance. Hence we can control its learning ability.
3. The algorithm finds, under general conditions a unique separating decision surface that provides the best out-of-sample performance
4. The architecture is feed-forward and is very amenable to parallel hardware implementation [4].

$Gini$SVM is a sparse form of multi-class large-margin probabilistic logistic regression. Using a cost function based

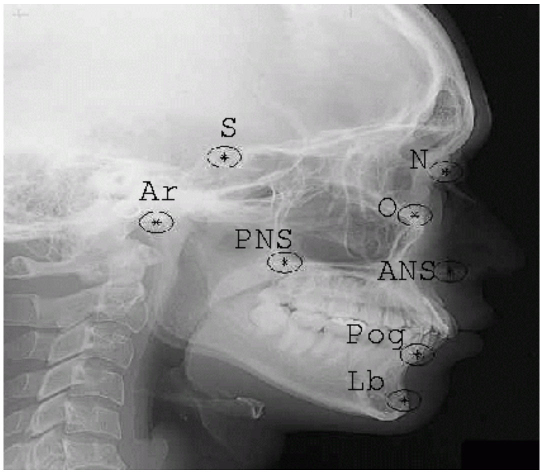

Fig. 1. *Eight landmark points used in cephalometric studies.*

on a quadratic ($Gini$) form of cross-entropy [7], training reduces to solving a quadratic programming problem under linear constraints [7]. Like any other kernel machine $Gini$SVM operate by mapping input vectors into very high dimensional feature space where a maximal margin hyperplane can be found that linearly separates the training data. The relevance of such high-dimensional feature processing is evident in this work where $Gini$SVM is able to easily extract non-linear features from training images, especially which are harder to identify using human cognition.

The paper is organized as follows. Section 2 describes the PPED feature extraction algorithm. Section 3 describes the $Gini$-SVM classifier architecture. Section 4 describes experiments performed using both the schemes on the task of cephalometric landmark detection. Section 5 provides conclusions and final remarks.

2. PROJECTED PRINCIPAL-EDGE DISTRIBUTION

PPED feature extraction tries to capture the information content of an image by modeling its edge distribution along different principal directions or orientations. For most general purposes four such directions suffice to model relevant discriminatory information. Details about PPED vector generation can be found in [2]. Here we enlist only the salient steps of feature extraction.

- Four principal directions, horizontal(H), vertical(V), clockwise 45^o(P45) and anti-clockwise 45^o (M45) are chosen along which edge detection has to be performed as shown in Figure 2. The edges along these directions are extracted using four 5x5 pixel edge detection filters [2]. A winner-take-all then selects the edge with the maximum intensity.

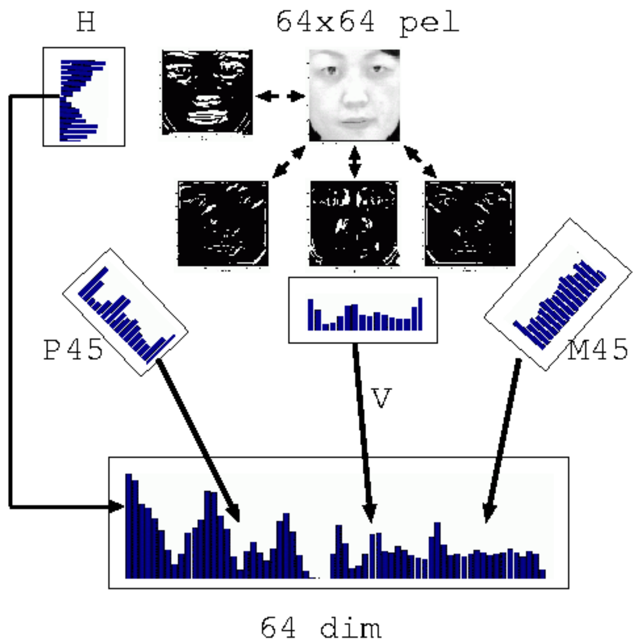

Fig. 2. *PPED feature generation algorithm.*

- The maximum edge intensity are compared with a threshold value to detect the presence of an edge. This eliminates effect of noise and illumination bias in the image. The threshold for an image is computed as the median of all the edge intensities present in the image details of which are provided in [2]. The detected image is then stored as a bitmap in one of the four orientations, shown in Figure 2.

- PPED coefficients are computed by projecting the bitmaps along directions orthogonal to the principal axes. Additional feature reduction is obtained by averaging along the coefficients, resulting in 16 features along each direction for a 64x64 pixel image [2].

- The 16 coefficients for each directions are then concatenated to form a composite 64 dimension PPED vector.

The PPED vectors, generated by the above procedure are used as a front-end for a $Gini$SVM classifier described below.

3. $GINI$-SUPPORT VECTOR MACHINE

In its basic form $Gini$SVM generates conditional probabilities $P(i|\mathbf{x})$ for a class/landmark $i, i = 1, .., K$ given input feature vector $\mathbf{x}$

$$P(i|\mathbf{x}) = exp(f_i(x)) / \sum_p exp(f_p(x)) \qquad (1)$$

As with SVMs, dot products in the expression for $f_i(\mathbf{x})$ in (1) convert into kernel expansions over the training data $\mathbf{x}[m], m = 1, .., M$ by transforming the data to feature space [6]

$$\begin{aligned} f_i(\mathbf{x}) &= \mathbf{w}_i.\mathbf{x} + b_i \\ &= \sum_m \lambda_i^m \ \mathbf{x}[m].\mathbf{x} + b_i \qquad (2) \\ &\xrightarrow{\Phi(\cdot)} \sum_m \lambda_i^m \ K(\mathbf{x}[m], \mathbf{x}) + b_i \end{aligned}$$

where $K(\cdot, \cdot)$ denotes any symmetric positive-definite kernel[1] that satisfies the Mercer condition, such as a Gaussian radial basis function or a polynomial spline [5].

The parameter λ_i^n are determined by solving linearly constrained quadratic programming problem [7] and parameter b_i is obtained as a lagrangian for the constraint (4)

$$H_g = \sum_i^M [\frac{1}{2} \sum_l^N \sum_m^N \lambda_i^l Q_{lm} \lambda_i^m + \gamma C \sum_m^N (y_i[m] - \lambda_i^m / C)^2] \qquad (3)$$

subject to constraints

$$\sum_m \lambda_i^m = 0 \qquad (4)$$

$$\sum_i \lambda_i^m = 0 \qquad (5)$$

$$\lambda_i^m \leq C y_i[m] \qquad (6)$$

Here $Q_{lm} = K(\mathbf{x}[m], \mathbf{x}[l])$ represents the kernel image matrix, and the additional parameters γ and C are obtained by tuning the performance of $Gini$SVM on a cross-validation set. The constrained optimization problem (3) can be solved by several standard quadratic programming techniques.

4. EXPERIMENTS AND RESULTS

For our experiments 130 x-ray head films (700x500 pixels) were taken from retention files at Department of Orthodontics, the Osaka University. A ten fold cross validation procedure was adopted to reliably evaluate the performance of the recognizer. 70 images were chosen for training, 20 for cross-validation experiments and 40 images were chosen to evaluate the performance of the trained recognizer. Such selections were repeated at random for ten times and the evaluation results were then averaged to obtain the final figure of merit. For all our experiments we chose a Gaussian kernel $K(\mathbf{x}, \mathbf{y}) = exp((\mathbf{x} - \mathbf{y})^T(\mathbf{x} - \mathbf{y}))$ due its superior convergence properties during training.

[1] $K(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x}).\Phi(\mathbf{y})$. The map $\Phi(\cdot)$ need not be computed explicitly, as it only appears in inner-product form.

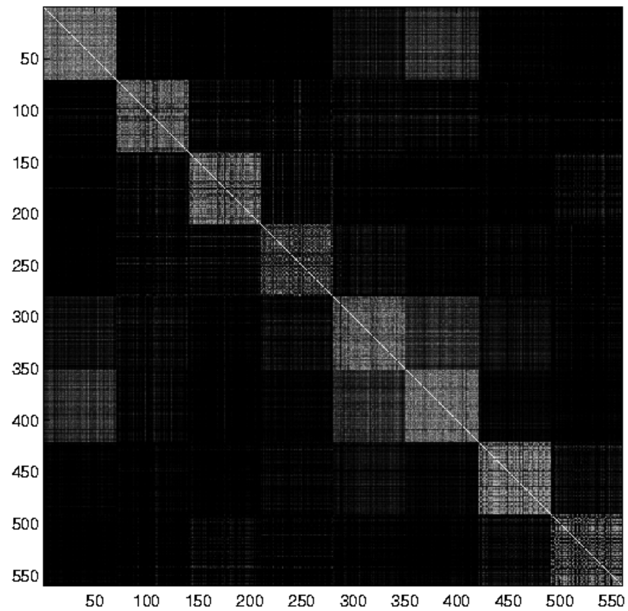

Fig. 3. *Image of a $Gini$SVM kernel map depicting the discriminatory power of PPED features for distinguishing between different cephalometric landmarks.*

The suitability of PPED vectors for $Gini$SVM can be observed through the training kernel image matrix shown in Figure 3. The distinct diagonal patterns in the image show that PPED feature vectors contain sufficient class discriminatory information which $Gini$SVM can directly exploit.

The second phase of the experiments included training data selection. For cephalometric identification, the landmarks have to distinguished from all the neighboring frames. In an ideal scenario all the background frames from the images can be included in training in which case the size of the set would easily exceed 10^8. Since SVM training involves modeling decision boundaries by classifying the worst examples correctly it would suffice to present only the background frames that serve as worst examples for landmark identification. Following procedure was adopted to extract the negative samples from all the training images.

- For each training image a template of PPED vectors corresponding to the landmark frames were extracted.
- The image was then scanned pixel by pixel and 64x64 pixel frame was extracted. PPED vectors for each frame was computed and its kernel distance $K(\mathbf{x}, \mathbf{y})$ was computed to extract the similarity metric to all landmark PPED vectors.
- Two PPED vectors for each landmark were chosen as negative candidates. The first one with the highest kernel score amongst all frames that were atleast

Table 1. Comparison of $Gini$SVM (SVM) recognition system with a Nearest-Neighbor (NN) classifier system

Landmarks	Tolerance(5mm)	Tolerance(1mm)
S (SVM)	91%	87%
S (NN)	84%	79%
N (SVM)	100%	95%
N (NN)	92%	87%
O (SVM)	100%	99%
O (NN)	95%	89%
Ar (SVM)	83%	79%
Ar (NN)	76%	67%
Pog (SVM)	100%	98%
Pog (NN)	85%	82%
Lb (SVM)	100%	99%
Lb (NN)	89%	81%
ANS (SVM)	98%	98%
ANS (NN)	92%	89%
PNS (SVM)	97%	96%
PNS (NN)	87%	81%

256x256 pixels away from the landmark. The second one with the highest kernel score amongst all the frames within 256x256 pixels of the landmark.

Prior knowledge about approximate locations of each landmark were encoded into the kernel by concatenating normalized frame centroid coordinates to each PPED vector. This procedure resulted in a training set consisting of 1680, 66 dimensional PPED vectors. $Gini$SVM was training on the training set and all the training parameters were tuned and optimized using the cross-validation set.

Evaluation of the recognizer was performed by scanning the test images pixel by pixel and extracting 64x64 frames, from which PPED vectors were generated and class conditional probabilities were computed using (1). The frame with the highest landmark class conditional probability out of all the frames in the image was chosen to be the location estimate for that landmark. Any location estimate that exceeded the true location estimate by a tolerance parameter was considered an error.

Table 1 compares the performance of the $Gini$SVM with a Nearest-Neighbor Classifier, for two tolerance values of the landmark identification points. The superior performance of $Gini$SVM can be directly attributed to the discriminant training of the landmarks and the background frames. The table also shows that given a larger tolerance value (5mm), more than 95% accuracy can be achieved for most landmarks using the SVM recognizer.

5. CONCLUSIONS

We demonstrated the robust performance of an image recognizer using PPED features with $Gini$SVM classifier for the task of Cephalometric Landmark Identification. Very accurate location estimation of the landmark can be obtained using the recognizer which is comparable to performance of expert dentists for a similar task. The utility of such a technique surpasses human performance especially when high dimensional non-linear features have to be evaluated for identification. Such a scenario occurs during identification of landmark *orbitale* 'O' and 'ANS', which the recognizer is able to identify with near perfect accuracy. Such recognizers can now be used for real-time image processing for which parallel architectures already exists [3] [4].

Acknowledgments

The authors would like to thank Prof. Kenji Takada and Chihiro Tanikawa, Department of Orthodontics, Faculty of Dentistry, Osaka Univ. for providing us the cephalometric head film samples and well as their expert knowledge of landmark identification.

6. REFERENCES

[1] Parthasarthy, S., Nugent, S.T., Gregson P.G and Fay D.F, "Automatic Landmarking of Cephalograms", *Computers and Biomedical Research*, vol. **22**, pp 248-269, 1989.

[2] Yagi, M. and Shibata, T. "A Human-Perception-like Image Recognition System based on PAP Vector Representation with Multi Resolution Concept," *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, vol. **1**, pp 1041-1048, 2001.

[3] Yamasaki, T., Yamamoto, K. and Shibata, T. "Analog Pattern Classifier with Flexible Matching Circuitry Based on Principal-Axis-Projection Vector Representation," *Proceedings of 27th European Solid-State Circuits Conference* , pp 212-215, 2001.

[4] Chakrabartty, S., Singh, G. and Cauwenberghs, G. "Hybrid Support vector Machine/Hidden Markov Model Approach for Continuous Speech recognition," *Proc. IEEE Midwest Symp. Circuits and Systems (MWSCAS'2000)*, Lansing, MI, Aug. 2000.

[5] Girosi, F., Jones, M. and Poggio, T. "Regularization Theory and Neural Networks Architectures," *Neural Computation*, vol. **7**, pp 219-269, 1995.

[6] Vapnik, V. *The Nature of Statistical Learning Theory,* New York: Springer-Verlag, 1995.

[7] Chakrabartty, S. and Cauwenberghs, G. "Forward Decoding Kernel Machines: A hybrid HMM/SVM Approach to Sequence Recognition," *IEEE Int. Conf. of Pattern Recognition: SVM workshop. (ICPR'2002)*, Niagra Falls, 2002.